

Notice of Allowability	Application No.	Applicant(s)	
	09/645,479	MANBER ET AL.	
	Examiner	Art Unit	

Adam L Basehoar
2178

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. This communication is responsive to 01/25/05.
2. The allowed claim(s) is/are 1-24 and 26-28.
3. The drawings filed on 24 August 2000 are accepted by the Examiner.
4. Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) All
 - b) Some*
 - c) None
 of the:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

* Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
 6. CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
 - (a) including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
 - 1) hereto or 2) to Paper No./Mail Date _____.
 - (b) including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).**
7. DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. Notice of References Cited (PTO-892)
2. Notice of Draftperson's Patent Drawing Review (PTO-948)
3. Information Disclosure Statements (PTO-1449 or PTO/SB/08),
Paper No./Mail Date _____
4. Examiner's Comment Regarding Requirement for Deposit
of Biological Material
5. Notice of Informal Patent Application (PTO-152)
6. Interview Summary (PTO-413),
Paper No./Mail Date _____.
7. Examiner's Amendment/Comment
8. Examiner's Statement of Reasons for Allowance
9. Other _____.

Art Unit: 2178

REASONS FOR ALLOWANCE AND EXAMINER'S AMENDMENT

1. This action is responsive to communications: The RCE filed 01/25/05.
2. Claim 25 has been cancelled as necessitated by Amendment.
3. The Rejection of claims 1-6, 8-10, 12-19, 21-23, and 26-28 under 35 U.S.C. 102(a) as being anticipated by Freivald et al (5,983,268 11/09/99) has been withdrawn as necessitated by Amendment.
4. The rejection of claims 7, 11, 20, and 24 under 35 U.S.C. 103(a) as being unpatentable over Freivald et al (5,983,268 11/09/99) in view of Ohashi (6,408,297 06/18/02) has been withdrawn as necessitated by Amendment.
5. The Specification has been amended by the examiner, which is discussed below in the Examiner's Amendment section.
6. Claims 1-24 and 26-28 are allowed as necessitated by Amendment and discussed below in the Reasons for Allowance section.

EXAMINER'S AMENDMENT

7. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

The Application has been amended as follows:

Replace the Abstract with: [Certain text is being removed from the Abstract because it purported merits or speculative applications of the invention which should not be included]

Systems and methods for analyzing HTML formatted web pages to automatically identify and extract desired information. A computer algorithm identifies and extracts different pieces of information from different web pages automatically after minimal manual setup. The algorithm automatically analyzes pages with different content if they have the same, or similar, formats. ~~The algorithm is fast and efficient and performs the extraction process quickly in real-time. The systems and methods are useful to build databases from unstructured web information.~~ The algorithm can be used as an agent that captures information about products, and compares prices or other characteristics. It can also be used to populate structured databases that, given the different pieces of information, can analyze products and their characteristics. And it can also be used for data mining application looking for patterns useful for marketing analyses, or other uses.

Replace the Specification with: [On Page 7 of the Specification, the sentence starting at the end of line 31 should be replaced with the following sentence because all hypertext links should be removed from the Specification]

One example of such an accessible server/database for which the invention is useful is the Yahoo! Shopping website at <http://shopping.yahoo.com>.

REASONS FOR ALLOWANCE

8. The following is an examiner's statement of reasons for allowance:

Regarding amended independent claim 1, none of the references, either singularly or in combination, teach or suggest to a person of ordinary skill in the art at the time of the invention the amended feature, "the generated symbols in the first string representing only HTML tags," in combination with previously presented claim limitations. The examiner notes that while using the tagged structure of HTML documents to compare and find similarities and differences between the documents is not considered a novel feature (See Examiner Referenced Patents), but in light of the combination of the features of the method steps of the independent claim the limitations of claim 1 are considered novel, and unobvious to a person of ordinary skill in the art at the time the invention was made in view of the prior art of record.

Regarding relevant prior art of record:

US-6,347,320 Christensen et al- Teaches a method for searching HTML documents on the Internet based on a search query generated from selecting specific HTML tags, wherein the specific HTML tags are used to compare the contents of the corresponding HTML tags in a database of HTML documents. Christensen does not teach generating a string of symbols representing only HTML tags, as Christensen relies on the contents of the corresponding tags to formulate the similarity between documents.

US-6,519,557 Emens et al- Teaches a system for identifying different language versions of the same structured format document. Emens accomplishes this by parsing two documents and constructing two hierarchical data structures to be compared (Fig. 3A). Emens does not

teach generating a first string of symbols representing only HTML tags and wherein an area of interest was identified only by a first portion of the first string.

US-6,526,410 Aoyama et al.- Teaches a document difference extraction method, which takes the logical meaning and structure of the document into consideration. Aoyama further teaches wherein a structured document extraction program compares only structured tags based on comparison criterion, and the contents of said tags are only compared when particular tags are coincident with each other to determine the difference of the two documents. Aoyama does not teach generating a first string of symbols representing only HTML tags and wherein an area of interest was identified only by a first portion of the first string.

US-6,502,112 Baisley- Teaches a system for identifying identical content between to compared XML documents based on a parsed and generated semantic graph (Fig. 2). Baisley does not teach generating a first string of symbols representing only HTML tags and wherein an area of interest was identified only by a first portion of the first string.

US-2005/0038785 Agrawal et al.-Teaches generating a numerical vector that represents the document structure based on the various elements of the document and the nature of their nesting. Agrawal then measures the similarity of the document structures to arrive at a measure of their similarity. While Agrawal does teach generating a first string of symbols representing HTML tags, Agrawal does not teach wherein a first area of user defined interest was identified by a first portion of the string of symbols.

As discussed above, neither of the references teach all the limitations of the claimed invention. In general the cited prior art describes various similar methods for comparing documents via their structure and/or their content to determine a certain measure of similarity.

The methods usually process the data in a generated hierarchical tree structure and not a generated string component representing only HTML tags that represents a unique number which was possibly represented in character sequence form (Specification: Page 8: lines 22-27).

Neither of the references teach wherein the first string generated for comparison identifies an area of interest by the first portion of the first string of symbols.

In regards to independent claims 12, 21, 22, and 26, the examiner notes that said claims are considered substantially similar to independent claim 1.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

US-2005/0038785	02-2005	Agrawal et al.
US-2004/0249824	12-2004	Brockway et al.
US-6,526,410	02-2003	Aoyama et al.
US-6,502,112	12-2002	Baisley, Donald
US-5,956,726	09-1999	Aoyama et al.
US-5,875,441	02-1999	Nakatsuyama, Hisashi
US-6,519,557	02-2003	Emens et al.

Art Unit: 2178

US-6,347,320

02-2002

Christensen et al.

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Adam L Basehoar whose telephone number is (571)-272-4121. The examiner can normally be reached on M-F: 7:00am - 4:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Steve Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ALB



STEPHEN HONG
SUPERVISORY PATENT EXAMINER